REMARKS

Claims 1-22 are pending in the application. Claims 1-22 stand rejected. Claim 2 was cancelled and replaced by added Claim 23. Claims 1, 3-4, 10-11, 17-18, and 20-21 were amended. Claims 1 and 3-23 remain in the application.

Claim 2 was inadvertently presented incorrectly in the previous amendment. Claim 2 has therefore been cancelled and replaced by Claim 23, which has the same language as original Claim 2. Claims dependent from Claim 2 were amended to depend from Claim 23.

The office action stated that Claims 1-2, 4-7 are rejected under 35 U.S.C. 102(a) as being anticipated by Qian et al., U.S. Patent No. 6,721,454 (hereafter "Qian 454"). From context, this rejection is also understood to apply to Claim 10. The rejection stated:

> 'As in claims 1 and 10, Qian et al. teaches a method and computer storage medium with instructions for obtaining unstructured video frames ("A video sequence 2 is input", Column 2, lines 64-65), generating segments from the shot boundaries based on the color dissimilarity between consecutive frames ("A color histogram technique may be used to detect the boundaries of the shots", Column 3, lines 42-43), extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61) for their visual dissimilarity and temporal relationship, and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16... events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).'

Claim 1 states:

> 1. A method for structuring video by probabilistic merging of video segments, said method comprising the steps of:
>
> a) obtaining a plurality of frames of unstructured video;
>
> b) generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames;

c) extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments; and

d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure.

Claim 1 is supported by the application as filed, notably the original claims and at page 11, line 30 to page 13, line 10.

The rejection indicates that "extracting a [feature] set by processing pairs of segments" is taught by '"the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61' of Qian 454. Segments are not frames. Claim 1 requires extracting a feature set by processing pairs of segments, the extracting generating features <u>of each pair of segments</u>: an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature. In Qian 454, global motion is not estimated between segments (shots), but rather between pairs of frames within a segment. Qian 454 states:

> "At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames <u>in a shot</u>." (Qian 454, col. 3, lines 59-61; <u>emphasis added</u>)

Qian 454 does teach a comparison of shot summaries, but does not disclose generating inter-segment features <u>of each pair of segments</u>. Qian states:

> "Referring to FIG. 10, a state diagram illustrates an animal hunt detection inference module. In this model inference module, a hunt event is inferred after detecting <u>three shots</u> containing hunt candidates (the video is tracking a fast moving animal) <u>which are followed by a shot</u> in which the video is no longer tracking a fast moving animal." (Qian 454, col. 11, lines 58-64; <u>emphasis added</u>)

Qian 454 also teaches against extracting inter-segment features by processing pairs of segments. In Qian 454, shots are compared in the form of summaries. Each of the individual shots, in Qian 454, are summarized with descriptors, such as "animal" and "tree", and the descriptors of different shots are

compared, but not in pairs. (Qian 454, col. 10, line 61 to col. 12, line 9) Qian teaches against comparisons between shots based upon "details":

> "The shot summaries provide a means of encapsulating the details of the feature and motion analysis performed at the first 4 and second 12 levels of the technique so that an event inference module in the third level 18 of the technique may be developed independent of the details in the first two levels. The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans. This facilitates video indexing, retrieval, and browsing in video databases and the development of algorithms to perform these activities." (Qian 454, col. 10, line 63 to col. 11, line 6; emphasis added)

In contrast, Claim 1 extracts a feature set by processing pairs of segments, generating inter-segment features.

Claim 1 requires generating segments by detecting shot boundaries and extracting a feature set by processing pairs of the segments. In the extracting, an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of segments are generated. The rejection indicates that the generating and extracting steps in Claim 1 are taught by '"A color histogram technique may be used to detect the boundaries of the shots", Column 3, lines 42-43' and '"the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61'. Unlike Claim 1, Qian 454 teaches that the "color histogram technique" and the global motion estimation both operate on frames not segments (shots). Qian 454 states:

> "A video sequence 2 is input to the first level 4 of the technique where it is decomposed into shots 6." (Qian 454, col. 2, lines 64-66, emphasis added)
>
> "At the first level of the technique 4, the boundaries of the constituent shots of the sequence are detected 6. A color histogram technique may be used to detect the boundaries of the shots." (Qian 454, col. 3, lines 40-43, emphasis added)
>
> "At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames in a shot." (Qian 454, col. 3, lines 59-61, emphasis added)

Qian 454 teaches event detection using summarization that encapsulates the details of the feature and motion analysis of each shot using descriptors.

(Qian 454, col. 10, line 63 to col. 11, line 8; col. 11, lines 51-55) In so doing, there is no generation of both an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each pair of segments. (Qian 454, col. 11, line 19 to col. 12, line 9; also see the descriptors discussed at Qian 454, col. 11, lines 7-18)

The rejection has also not addressed how use in merging, of features including a color dissimilarity feature, would be compatible with the disclosed summarization of Qian 454. The disclosed shot descriptors in Qian 454, relate to objects in a frame and their relationships--both spatial and temporal. Such objects and relationships are very unlike color dissimilarity between segments. (Qian 454, col. 11, lines 7-18)

Claim 1 further requires a step merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set. The rejection indicates that "merging the video segments by applying a probalistic analysis to the extracted set to represent the video structure" is taught by '"each shot is summarized 16 ... events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8'. Where in Qian 454 is there a teaching or suggestion of a probabilistic analysis of inter-segment features of pairs of segments? The analysis of the model inference module of the hunt event (Qian 454, col. 11, lines 58-64; quoted above) considers if descriptors provide a "true" value in three successive shots. (Qian 454, col. 11, line 58 to col. 12, line 3) U.S. Patent No. 6,616,529 (Qian 529) discloses an application of a Bayesian analysis to detected semantic events. (Qian 529, col. 2, lines 56-57) Assuming for the sake of argument that Qian 454 and Qian 529 could be combined, a Bayesian analysis would apparently replace the "true" values of Qian 454, col. 11, line 58 to col. 12, line 3 with probabilities. Such a combination would still only teach application of a Bayesian model to shot descriptors or semantic events and not a probabilistic analysis of inter-segment color dissimilarity and temporal relationship features of pairs of segments.

Claims 23 (which replaces cancelled Claim 2) and 4-7 are allowable as depending from Claim 1 and as follows.

Claim 4 states:

> 4. The method as claimed in claim 23 further including the step of morphologically transforming the thresholded

difference signal with a pair of structuring elements that eliminate the presence of multiple adjacent shot boundaries.

The rejection of Claim 4 does not address the claim as written. The rejection states:

> "As in Claim 4, Qian et al. teaches morphologically transforming the thresholded difference signal with a pair of structuring elements to eliminate the presence of multiple adjacent shot boundaries ("When the difference between the histograms of two frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different", Column 3, lines 45-48)." (emphasis added)

The office action also comments in the Response to Arguments:

> "In response to the arguments regarding claim 4, Qian does teach detection of shot boundaries admitted by the applicant on page 9, line 29."

Claim 4 does not specify detecting shot boundaries. Claim 4 requires morphologically transforming the thresholded difference signal with a pair of structuring elements that eliminate the presence of multiple adjacent shot boundaries. Where does Qian 454 teach or suggest morphologically transforming the thresholded difference signal with a pair of structuring elements that eliminate the presence of multiple adjacent shot boundaries? It is noted that Qian 454, in contrast, teaches providing additional shot boundaries by forcing or inserting:

> "In addition to the shot boundaries detected in the video sequence, shot boundaries may be forced or inserted into the sequence whenever the global motion of the content changes. As a result, the global motion is relatively homogeneous between the boundaries of a shot. In addition, shot boundaries may be forced after a specific number of frames (e.g., every 200 frames) to reduce the likelihood of missing important events within extended shots." (Qian 454, col. 3, lines 51-58)

The rejection stated in relation to Claim 5:

> "As in Claim 5, Qian et al. teaches computing a mean color histogram for each segment and a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments (Column 3, lines 42-50 and Figure 5)."

Claim 5 states:

> 5. The method as claimed in claim 1 wherein the processing of pairs of segments for visual dissimilarity in step c) comprises the steps of computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments.

The cited portion of Qian 454 relates to determining the difference between histograms of frames to detect <u>shot boundaries</u>. Qian 454 states (quoting at greater length):

> "A video sequence 2 is input to the first level 4 of the technique where it is decomposed into shots 6." (Qian 454, col. 2, lines 64-66)
>
> "At the first level of the technique 4, the boundaries of the constituent shots of the sequence are detected 6. A color histogram technique may be used <u>to detect the boundaries of the shots</u>. The difference between the histograms of two frames indicates a difference in the content of those frames. When the difference between the histograms for successive frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be used to detect the shot boundaries." (Qian 454, col. 3, lines 40-50)

Claim 5 describes a feature of step c) of Claim 1, in which pairs of segments are processed. The segments are products of step b) of Claim 1: generating video segments by detecting shot boundaries. The cited portions of Qian 454 discuss use of a color histogram technique in relation to detecting shot boundaries. This, arguably, relates to step b) of Claim 1. The cited portions of Qian 454 do not teach or suggest use of a color histogram technique in processing earlier-detected segments for visual dissimilarity. As earlier noted, Qian teaches against use of details in the first two levels for comparisons of shots. Claim 5 also requires:

> computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments.

Even if Qian related to the appropriate step of the claimed method, the portion of Qian cited in relation to Claim 5 teaches taking a difference between histograms of two <u>frames</u>. Qian has no teaching of taking a <u>mean color histogram</u> for each of two segments and then taking a difference between the two mean histograms.

The rejection stated in relation to Claim 6:

> 'As in Claim 6, Qian et al. teaches processing pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration between pairs of segments ("each shot is summarized 16. . . events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).'

Claim 6 states:

> 6. The method as claimed in claim 1 wherein the processing of pairs of segments for their temporal relationship in step c) comprises the processing of pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration between pairs of segments.

Qian 454 teaches summarization that encapsulates the details of the feature and motion analysis of each shot using descriptors. (Qian 454, col. 10, line 63 to col. 11, line 8) The domain specific event inference model uses the descriptors. (Qian 454, col. 11, lines 51-55) Events are inferred by matching the occurrence of objects and their spatial and temporal relationships detected in each of the shots. (Qian 454, col. 12, lines 6-7; generally see col. 11, line 58 to col. 12, line 9) Examples of shot descriptors are provided:

> 'In general, shot descriptors used in the shot summary include object, spatial, and temporal descriptors. The object descriptors indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc. The spatial descriptors represent location and size information related to objects and the spatial relations between objects in terms of spatial prepositions, such as "inside", "next to", "on top of", etc. <u>Temporal descriptors represent motion information related to objects and the temporal relations between them.</u> These may be expressed in temporal prepositions, such as, "while", "before", "after," etc.' (Qian 454, col. 11, lines 7-18; <u>emphasis added</u>)

Qian 454 does not teach descriptors of temporal separation between pairs of segments and/or for accumulated temporal duration between pairs of segments. In Qian 454, temporal descriptors represent motion information related to objects in a segment and the temporal relations between those

objects in that segment. This is unlike Claim 6, which requires processing pairs of segments for a <u>temporal separation between</u> pairs of segments and for an <u>accumulated temporal duration between</u> pairs of segments.

The rejection stated in relation to Claim 7:

> "As in Claim 7, Qian et al. teaches generating parametric mixture models (summaries created by shot summarization 16, Figure 1) to represent class-conditional densities of inter-segment features (based on temporal information and color analysis, See Claim 1 rejection supra) of the feature set and applying the merging criterion to the parametric mixture models (event inference 20/detected events 22, Figure 1)."

Claim 7 states:

> 7. The method as claimed in claim 1 wherein step d) comprises the steps of:
>
> generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set; and
>
> applying the merging criterion to the parametric mixture models.

Claim 7 requires generating "parametric mixture models" that are defined by the specification and and usage in the art as types of statistical models. (See application page 4, lines 25-30; page 13, lines 14-29; also see U.S. Patent No. 5,710,833.) The rejections's "summaries created by shot summarization" are not statistical models. Qian 454 teaches summaries, in which shot descriptors are described as indicating as to a particular shot: "the existence of certain objects", "location and size information related to objects and the spatial relations between objects", and "motion information related to objects and the temporal relations between them". (Qian 454, col. 11, lines 9, 11-13, and 15-16; see also the above discussion of summarization.) One can attempt to combine the example shot descriptors in Qian 454 in this manner. For example, one could say --animal inside tree while second animal on top of rock--. How is this form of summarization compatible with a statistical model, such as a parametric mixture model? In Claim 7, the parametric mixture models are generated to represent class-conditional densities of <u>inter-segment</u> features. This contrasts with the shot descriptors for each segment

taught by Qian 454 and discussed above in relation to Claim 6. (Also see Qian 454, col. 10, lines 61-62: "Each shot ... is summarized").

Claim 10 is supported and allowable on the grounds discussed above in relation to Claim 1.

Claim 3 stands rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., U.S. Patent No. 6,721,454. Claim 3 is allowable as depending from Claim 1.

Claim 8 stands rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., U.S. Patent No. 6,721,454. The rejection stated:

> "In accordance with Claims 8 and 15, it is notoriously well known that queues are used to implement hierarchical displays. The examiner takes official notice of this teaching. It would be obvious to one of ordinary skill in the art to combine the use of the organizing video segements into hierarchies with a queue implementation."

Claim 8 is allowable as depending from Claim 1 and as follows. Claim 8 states:

> 8. The method as claimed in claim 7 wherein step d) is performed in a hierarchical queue and comprises the steps of:
>
> initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;
>
> depleting the queue by merging the segments if the merging criterion is met; and
>
> updating the model of the merged segment and then updating the queue based upon the updated model.

The rejection argues that it is notoriously well known that queues are used to implement hierarchical displays. This statement addresses only one phrase of Claim 8: "performed in a hierarchical queue" and does not teach or suggest the steps of:

> initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;
>
> depleting the queue by merging the segments if the merging criterion is met; and

updating the model of the merged segment and then updating the queue based upon the updated model.

The rejection also does not teach or suggest perforance of step d (that is, merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure) in a hierarchical que.

Claims 9, 11-15 and 16-22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., U.S. Patent No. 6,721,454 ("Qian 454"), and further in view of Qian et al., U.S. Patent No. 6,616,529 (hereafter "Qian 529"). The rejection states:

'As in Claims 9, 11, 17-18 and 20, US Patent 6721454 teaches a method and computer storage medium with instructions for obtaining unstructured video frames ("A video sequence 2 is input", Column 2, lines 64-65), generating segments from the shot boundaries based on the color dissimilarity between consecutive frames ("A color histogram technique may be used to detect the boundaries of the shots", Column 3, lines 42-43), extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61) for their dissimilarity and temporal relationship, merging adjacent video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8), and generating a parametric mixure model of the inter-segment features ("In this model inference module, a hunt event is inferred after detecting three shots containing hunt candidates", Column 11, lines 60-62). While US Patent 6721454 teaches the segmentation due to color dissimilarity, extraction due to visual dissimilarity and temporal relationships, merging probabilistic analysis and generation of a parametric mixture model, they fail to show the probabilistic analysis to be a Bayesian analysis applied to the parametric mixture model, and representing the merging sequence in a hierarchical tree structure as recited in the claims. US Patent 6616529 teaches a video segmentation method similar to that of US Patent 6721454. In addition, US Patent

6616529 further teaches the probabilistic analysis to be a Bayesian analysis applied to the parametric mixture model (Figure 3 and corresponding text in Columns 4-5), and representing the merging sequence in a hierarchical tree structure (Figures 2a-2g and corresponding text). It would be obvious to one of ordinary skill in the art, having the teachings of US Patent 6721454 and US Patent 6616529 before him at the time the invention was made, to modify the segmentation with color dissimilarity and temporal relationships with a parametric mixture model taught by US Patent 6721454 to include the construction of hierarchy according to probabilistic merging with Bayesian analysis of US Patent 6616529, in order to obtain a hierarchical representation of the frames grouped by color dissimilarity and temporal relationships according to Bayesian probability methods of analysis. One would have been motivated to make such a combination because a visual representation of the segmented video would have been obtained, as taught by US Patent 6616529 (Column 2, lines 24-55).'

Claim 9 is allowable as depending from Claim 1.

Claim 11 is supported and allowable as discussed above in relation to Claim 1.

Claims 12-16 are allowable as depending from Claim 11. Claims 12-15 are also allowable on the same basis as Claims 5-8, respectively.

Claims 17-18 are supported and allowable on grounds discussed above in relation to Claim 1.

Claim 19 is allowable as depending from Claim 18.

Claims 20-21 are supported and allowable on grounds discussed above in relation to Claim 1.

Claim 22 is allowable as depending from Claim 21.

It is believed that these changes now make the claims clear and definite and, if there are any problems with these changes, Applicants' attorney would appreciate a telephone call.

In view of the foregoing, it is believed none of the references, taken singly or in combination, disclose the claimed invention. Accordingly, this application is believed to be in condition for allowance, the notice of which is respectfully requested.

Respectfully submitted,

Attorney for Applicant(s)
Registration No. 30,700

Robert Luke Walker/amb
Rochester, NY 14650
Telephone: (585) 588-2739
Facsimile: (585) 477-1148

If the Examiner is unable to reach the Applicant(s) Attorney at the telephone number provided, the Examiner is requested to communicate with Eastman Kodak Company Patent Operations at (585) 477-4656.